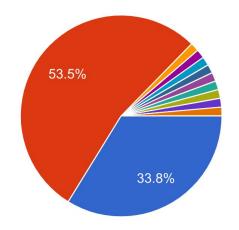
Data Storage and Access (Software)

Input data format for analysis

Completely dominated by ROOT-based formats

What data storage format do you run most of your own code on? 71 responses

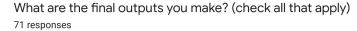


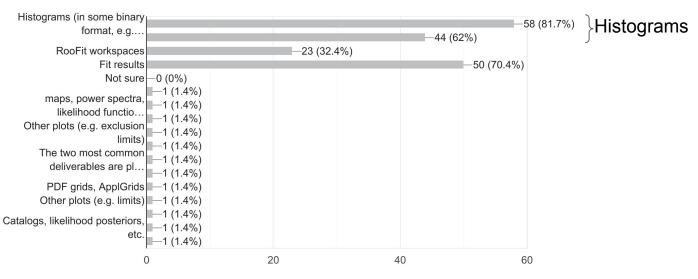
- "plain ROOT" (flat TTrees, etc.)
- ROOT-based with experiment-specific...
- Experiment-specific, non-ROOT binar...
- Text-based (e.g. CSV files)
- HDF5
- FITS
- Parquet
- SQL selections from database



Final outputs

Mostly histograms + final fit results. About a third produce RooFit workspaces.

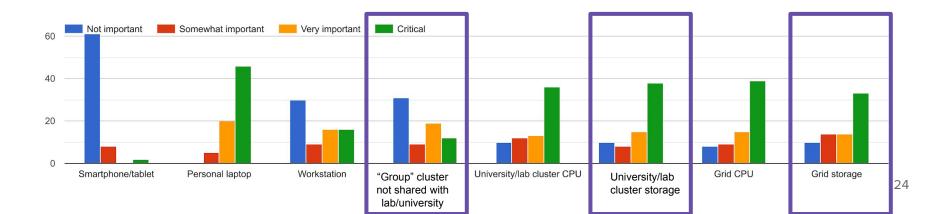




Where are the data?

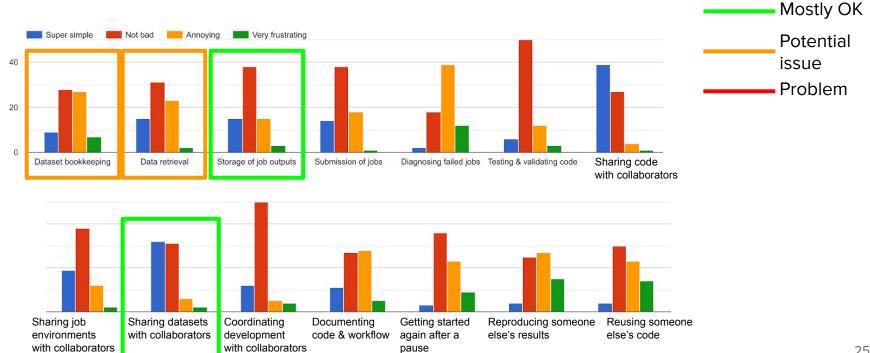
People dominantly work from laptops
Larger scale (university/lab-level or above) storage more important than "group"
computing

Rank the importance of the following computing resources to your analysis.



What's frustrating?

Rate how easy it is to deal with the following elements in your analysis.



Convener Questions

- Are we happy with ROOT as a standard data format? Are there techniques that would be enabled with different formats?
- Do we have adequate tools to bookkeep analysis data (as opposed to e.g. production datasets)?
- Would analysis outputs other than histograms be useful? Are there barriers to producing them?
- Are there mismatches of where data are and where people want to do compute? (e.g. Grid storage element vs user laptop)